EXTRACT

[0008]

[Means for Solving the Problem]

An input/output control system in a computer system according to the present invention having a check point roll back mechanism which performs a process while periodically check points and which returns the state of the system to a state at the time of the latest picked check point when a fault occurs in the system to re-execute the process and an input/output module which writes and reads data in/from a secondary storage device, includes an invalid block management module which manages an address conversion map which allocates at least one physical block to each of logic blocks constituting the secondary storage device and which allocates, to the logic blocks, a pointer for a physical block having a logic block image at the time of the latest check point and a pointer for a logic block having the latest logic block image and a physical block which do not have valid data on the secondary storage device, and is characterized in that, when the input/output module accepts a write request for the secondary storage device, data is written in a physical block different from the latest check point image with respect to the logic blocks, and the physical block is registered in the address conversion table as the latest logic block image, when the input/output module accepts a read request for the secondary storage device, data is read from a physical block having the latest logic block image with respect to the logic blocks, in a check point process,

with respect to the logic blocks on the address conversion map, the physical block having the latest logic block image is set as a physical block having a check point image for the logic blocks and the latest image, in a recovery operation, with respect to the logic blocks on the address conversion map, a physical block having a logic block at the time of a check point as a physical block having a check point image for the logic blocks and the latest image.

[0009]

The input/output control system is characterized in that times at which data are finally written in the logic blocks are further recorded on the address conversion map, when the input/output module accepts a write request for the secondary storage device, data is written in a physical block different from the latest check point image with respect to the logic blocks, the physical block is registered as the latest logic block image, time at which a write process is performed to the physical block is registered on the address conversion map, when the input/output module accepts a read request from the secondary storage device, data is read from a physical block having the latest logic block image with respect to the logic block, in a check point process, the present time is stored as a check point time stamp, in a recovery operation, with respect to the logic blocks on the address conversion map, when a pointer for the latest physical block having a time stamp newer than the check point time stamp is registered, the latest physical block is given to an invalid block management module

and released, and a pointer for a physical block at the time of the check point is registered on the address conversion map as a pointer for the latest physical block.

[0010]

The input/output control system further includes a difference recording mechanism which records, with respect to a logic block subjected to a write process after the time of the latest check point, a combination of the logic block and a physical block allocated to the logic block before the write process, and is characterized in that, when the input/output module accepts a write request for the secondary storage device, the logic block at this time and mapping on an address conversion map of physical blocks on the difference recording mechanism, a physical block newly secured from the invalid block management module on the address conversion map as a physical block for the logic block, a write process is performed to the physical block, in a check point process, all physical blocks recorded on the difference recording mechanism are released for the invalid block management module, all combinations of logic blocks and physical blocks recorded on the difference recording mechanism are deregistered, in a recovery operation, with respect to all the logic blocks recorded on the difference recording mechanism, physical blocks mapped to the logic blocks by the address conversion map are given to the invalid block management module and released, and the combinations of the logic blocks and the physical blocks registered on the difference recording

mechanism are written back to restore the address conversion map at the time of the latest check point.

[0011]

The input/output control system is characterized in that, when the input/output module accepts a write request for the secondary storage device, if a designated logic block is not registered on the difference recording mechanism, mapping of the logic block at this time and a physical block on the address conversion map on the difference recording mechanism, after a physical block newly secured from the invalid block management module is registered on the address conversion map as a physical block for the logic block, a write process is performed to the physical block, if the designated logic block is registered on the difference recording mechanism, data is written in a physical block mapped to the logic block by the address conversion map, in a check point process, all the physical blocks recorded on the difference recording mechanism are released for the invalid block management module, all combinations of logic blocks and physical blocks recorded on the difference recording mechanism are deregistered, in a recovery operation, with respect to the all logic blocks recorded on the difference recording mechanism, physical blocks mapped to the logic blocks by the address conversion map are given to the invalid block management module and released, and the combinations of the logic blocks and the physical blocks registered on the difference recording mechanism are written back to restore the contents of the

address conversion map at the time of the latest check point.

[0012]

According to the above configuration, a computer system using a check point roll back mechanism, since an input/output operation for the secondary storage device can be carried out, the performance of the system is improved. Furthermore, since the number of redundant physical blocks on the secondary storage device can be reduced, the capacity efficiency of the secondary storage device can also be improved.

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-047891
(43)Date of publication of application : 18.02.2000

(51)Int.CI. G06F 11/14
G06F 3/06
G06F 12/00
G06F 12/10

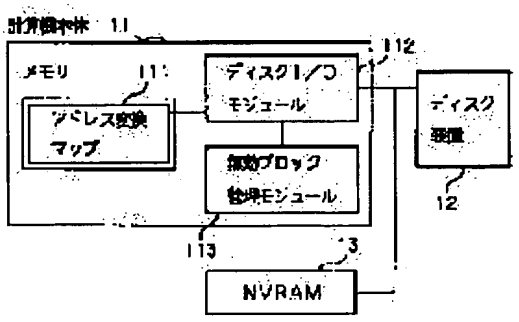
(21)Application number : 11-048453
(22)Date of filing : 25.02.1999

(71)Applicant : TOSHIBA CORP
(72)Inventor : SHIMIZU KUNIYASU

(30)Priority
Priority number : 10150041 Priority date : 29.05.1998 Priority country : JP

(54) DEVICE AND METHOD FOR CONTROLLING INPUT/OUTPUT FOR COMPUTER SYSTEM AND STORAGE MEDIUM STORING THE PROGRAMMED METHOD

(57)Abstract:
PROBLEM TO BE SOLVED: To easily realize the capacity efficiency improvement of a secondary storage device.
SOLUTION: This device is provided with an address conversion map 111 assigning at least one logical lock to the logic block of a disk device 12 and assigning a pointer to a physical block with a logic block image at the time point of the newest checking point and a pointer to a physical block with a newest block image to each block, and an ineffective block managing module 113 managing a physical block without effective data on the disk device. Then, in a check point processing, the physical block with the newest logic block image in each logic block on the conversion map is set to be a physical block with a check point image with respect to the logical block and a newest image, and the physical block with the logic block image at the time point of a checking point in each logic block on the conversion map at the time of recovery is set to be a physical block with a check point image with respect to the logic block and the newest image.



LEGAL STATUS

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－47891
（P2000－47891A）
(43)公開日　平成12年２月18日(2000.2.18)

(51)Int.Cl.⁷	識別記号	FI		テーマコード（参考）
G06F 11/14	310	G06F 11/14	310B	
3/06	302	3/06	302A	
12/00	501	12/00	501H	
12/10		12/10	E	

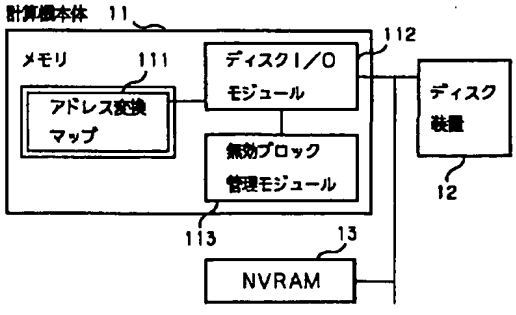
審査請求　未請求　請求項の数22　ＯＬ　（全 21 頁）

(21)出願番号　特願平11－48453
(22)出願日　平成11年２月25日(1999.2.25)
(31)優先権主張番号　特願平10－150041
(32)優先日　平成10年５月29日(1998.5.29)
(33)優先権主張国　日本（ＪＰ）

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者　清水　邦保
東京都青梅市末広町２丁目９番地　株式会社東芝青梅工場内
(74)代理人　100081732
弁理士　大胡　典夫　（外１名）

(54)【発明の名称】　計算機システムにおける入出力制御装置及び同システムにおける入出力制御方法並びに同方法がプログラムされ記憶された記憶媒体

(57)【要約】
【課題】　二次記憶装置の容量効率改善を容易に実現すること。
【解決手段】　ディスク装置１２の論理ブロックに少なくとも一つの物理ブロックを割当て、各ブロックに最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新論理ブロックイメージを持つ物理ブロックへのポインタを割当てるアドレス変換マップ１１１、ディスク装置上で有効データを持たない物理ブロックを管理する無効ブロック管理モジュール１１３を有し、チェックポイント処理では上記変換マップ上の各論理ブロックに最新論理ブロックイメージを持つ物理ブロックを該論理ブロックに対するチェックポイントイメージかつ最新イメージを持つ物理ブロックとし、リカバリ時に上記変換マップ上の各論理ブロックにチェックポイント時点の論理ブロックイメージを持つ物理ブロックを該論理ブロックに対するチェックポイントイメージかつ最新イメージを持つ物理ブロックとする。



1

【特許請求の範囲】

【請求項１】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも一つの物理ブロックを割り当て、各論理ブロックに対して、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ論理ブロックへのポインタを割り当てるアドレス変換マップと、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置に対する読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、

チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージを持つ物理ブロックとし、

リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、チェックポイント時点の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージをもつ物理ブロックとすることを特徴とする計算機システムにおける入出力制御方法。

【請求項２】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、更にそれぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録するアドレス変換マップと、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして登録すると共にこの物理ブロックに書き込み処理を行なった時刻をアドレス変換マップに記録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、

チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、

リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、上記チェックポイントタイムスタンプよりも新しいタイムスタンプを持つ最新の物理ブロックへのポインタが登録されている場合は最新の物理ブロックを上記無効ブロック管理モジュールに渡して解放し、チェックポイント時点の物理ブロックへのポインタを最新の物理ブロックへのポインタとして上記アドレス変換マップに登録することを特徴とする計算機システムにおける入出力制御方法。

【請求項３】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関し、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、この時点のこの論理ブロックと物理ブロックのアドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行い、

チェックポイント処理では、上記差分記録機構に記録した全ての物理ブロックを無効ブロック管理モジュールに

解放し、上記差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、
リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元することを特徴とする計算機システムにおける入出力制御方法。

【請求項４】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、
最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関し、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールを有し、
上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときに、指定された論理ブロックが上記差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行い、指定された論理ブロックが上記差分記録機構に登録されていれば上記アドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、
チェックポイント処理では、上記差分記録機構に記録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、
リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップの内容を復元することを特徴とする計算機システムにおける入出力制御方法。

【請求項５】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、
上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、リカバリによって最新チェックポイント時点の状態に戻るメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、上記二次記憶装置とメモリの間でアドレス変換マップのページングを行なうアドレス変換マップページング機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとをもち、
上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構により上記メモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックの最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、
上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、
チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージと最新のイメージを持つ物理ブロックとするようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項６】　定期的にチェックポイントを採取しなが

ら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、リカバリによって最新チェックポイント時点の状態に戻らないメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なうアドレス変換マップページング機構と、最新チェックポイント時点以降のメモリ上の上記アドレス変換マップの変更履歴を保存するアドレスマップ差分テーブルと、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によってメモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックの最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合には、上記アドレスマップページング機構により上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージを有する物理ブロックからデータを読み出し、

チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージと最新のイメージを持つ物理ブロックとし、上記アドレスマップ差分テーブルに登録されたエントリを抹消し、

リカバリ時には、上記アドレスマップ差分テーブル登録された各エントリを上記アドレス変換テーブルに書き戻すことを特徴とする計算機システムにおける入出力制御方法。

【請求項７】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、それぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録し、リカバリによって最新チェックポイント時点に戻るメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なうアドレス変換マップページング機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップがメモリ上に存在しない場合、上記アドレス変換マップページング機構によりメモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとすること、およびこの物理ブロックに書き込み処理を行なった時刻を上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、

チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存するようにしたこと特徴とする計算機システムにおける入出力制御方法。

【請求項８】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、それぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録し、リカバリによって最新チェックポイント時点の状態に戻らないメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、二次記憶装置とメモリの間でアドレス変換マップのページングを行なうアドレス変換マップページング機構と、最新チェックポイント時点以降のメモリ上のアドレス変換マップの変更履歴を保存するアドレスマップ差分テーブルと、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとすること、およびこの物理ブロックに書き込み処理を行なった時刻を上記アドレス変換マップに登録し、

上記入出力モジュールが二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、

チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、上記アドレスマップ差分テーブルに登録されたエントリを抹消し、

リカバリ時には、上記アドレスマップ差分テーブルに登録された各エントリを上記アドレスマップ差分テーブルに書き戻すようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項９】　チェックポイント/ロールバック機構を有し、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうリカバリにより最新のチェックポイント時点の状態に戻らないメモリ領域、および上記二次記憶装置上にアドレス変換マップを有し、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理をおこなった論理ブロックに関してこの論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なうアドレス変換マップページング機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、この時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行ない、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み

出し、
チェックポイント処理では、上記差分記録機構に記録したすべての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に記録してあるすべての論理ブロックと物理ブロックの組の登録を抹消し、
リカバリ時には、上記差分記録機構に記録してあるすべての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元するようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項１０】　チェックポイント/ロールバック機構を有し、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうリカバリによりて最新チェックポイント時点の状態に戻らないメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップをもち、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なうような入出力モジュールをもつ計算機システムにおいて、
上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なう上記アドレス変換マップページング機構をもち、最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関して、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、
上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次回のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、指定された論理ブロックが上記差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行ない、指定された論理ブロックが上記差分記録機構に登録されていれば上記アドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、
上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップがメモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に次回のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、
チェックポイント処理では、上記差分記録機構に記録したすべての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に記録してあるすべての論理ブロックと物理ブロックの組の登録を抹消し、
リカバリ時には、上記差分記録機構に記録してあるすべての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元するようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項１１】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールが記憶装置に割り付けられ格納される計算機システムにおいて、
上記二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも一つの物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、ならびに最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当てるアドレス変換マップが割り付けられ保持される記憶手段と、
上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理手段と、
上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、上記入出力モジュールが上記二次記憶装置に対する読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブ

ロックからデータを読み出し、チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージを持つ物理ブロックとし、リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、チェックポイント時点の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージをもつ物理ブロックとする入出力制御手段とを具備することを特徴とする計算機システムにおける入出力制御装置。

【請求項１２】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、更にそれぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録するアドレス変換マップが割り付けられ保持される記憶手段と、

二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理手段と、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして登録すると共にこの物理ブロックに書き込み処理を行なった時刻をアドレス変換マップに登録し、上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、リカバリ時には、アドレス変換マップ上の各論理ブロックに対して、チェックポイントタイムスタンプよりも新しいタイムスタンプを持つ最新の物理ブロックへのポインタが登録されている場合は最新の物理ブロックを上記無効ブロック管理モジュールに渡して解放し、チェックポイント時点の物理ブロックへのポインタを最新の物理ブロックへのポインタとして上記アドレス変換マップに登録する入出力制御手段とを具備することを計算機システムにおける入出力制御装置。

【請求項１３】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関してこの論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組が記録される差分記録機構と、

上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理手段と、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、この時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行い、チェックポイント処理では、上記差分記録機構に記録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元する入出力制御手段とを具備することを特徴とする計算機システムにおける入出力制御装置。

【請求項１４】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関してこの論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組が記録される差分記録手段と、

二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理機構と、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときに、指定された論理ブロックが上

記差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行い、もし指定された論理ブロックが上記差分記録機構に登録されていれば上記アドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、チェックポイント処理では、上記差分記録機構に記録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理機構に渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップの内容を復元する入出力制御手段とを具備することを特徴とする計算機システムにおける入出力制御装置。

【請求項１５】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも１つの物理ブロックを割り当て、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージを持つ物理ブロックへのポインタを割り当てるアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、その論理アドレスに対して最新の論理ブロックイメージが存在するか否かチェックさせる機能と、上記アドレス変換マップを参照することにより、指定された論理ブロックに対して最新の論理ブロックイメージが登録されている場合には、最新の論理ブロックイメージとして登録されている物理ブロックに書き込み処理を行なわせる機能と、指定された論理ブロックに対して最新の論理ブロックイメージが割り当てられていない場合には、無効ブロック管理モジュールから新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの最新のイメージを持つ有効物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項１６】　チェックポイント処理において、アドレス変換マップに登録されている全ての最新チェックポイント時点の論理ブロックイメージとなる物理アドレスエントリを前記無効ブロック管理モジュールに渡すことにより解放し、全ての最新チェックポイント時点の論理ブロックイメージとなる物理アドリスエントリを物理ブロックが割り当てられていないことを示す特定の値に設定させる機能と、リカバリ処理において、上記アドレス変換マップを参照することにより最新チェックポイント時点の論理ブロックイメージとなる物理アドレスエントリが上記特定の値でない全ての論理ブロックに対して対応する有効物理ブロックを上記無効ブロック管理モジュールに渡すことで解放し、最新チェックポイント時点の論理ブロックイメージとなる物理ブロックアドレスエントリの値を最新の論理ブロックイメージとなる物理ブロックアドレスエントリに移動し、最新チェックポイント時点の論理ブロックイメージとなる物理ブロックアドレスエントリに対し上記特定の値を設定させる機能と、システムシャットダウン時には、アドレス変換マップの内容を不揮発性記憶領域に保存し、次回ブート時にはアドレス変換マップの内容をシャットダウン完了時点の状態に復元させる機能とがプログラムされ記録されるコンピュータ読み取り可能な請求項１５記載の記憶媒体。

【請求項１７】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも１つの物理ブロックを割り当て、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージを持つ物理ブロックへのポインタ、タイムスタンプを割り当てるアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、最新の論理ブロックイメージとなるブロックのタイムスタンプが最新のチェックポイントタイムスタンプより古いか否かをチェックさせる機能と、上記アドレス変換マップの内容を参照することにより、指定された論理ブロックに対して有効な物理ブロックのタイムスタンプがチェックポイントタイムスタンプよりも古くない場合には、現在有効な物理ブロックとして登録されている物理ブロックに書き込み行なわせる機能と、指定された論理ブロックに対して最新の論理ブロックイメージである物理ブロックのタイムスタンプがチェックポイントタイムスタンプよりも古い場合、更に最新チェックポイント時点の論理ブロックイメージであるアドレスが存在するか否かをチェックさせる機能と、指定された論理ブロッ

クの最新チェックポイント時点の論理ブロックイメージである物理ブロックアドレスが上記特定の値でなければ、この物理ブロックを無効ブロック管理モジュールに渡すことで解放し、指定された論理ブロックに対して、最新の論理ブロックイメージである現在有効な物理ブロックのアドレスを最新チェックポイント時点の論理ブロックイメージの物理ブロックアドレスとして登録させる機能と、上記無効ブロック管理モジュールから新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項１８】　チェックポイント処理において、現在の時刻をチェックポイントタイムスタンプとして登録させる機能と、リカバリ処理において、アドレス変換マップ上の全ての論理ブロックに対して、その論理ブロックに対応する有効物理ブロックのタイムスタンプがチェックポイントタイムスタンプよりも新しい場合は、有効物理ブロックを上記無効ブロック管理モジュールに渡すことで解放し、最新チェックポイント時点の論理ブロックイメージである物理ブロックアドレスエントリの値を最新の論理ブロックイメージである物理ブロックアドレスに移動し、最新チェックポイント時点の論理ブロックイメージである物理ブロックアドレスエントリの値を上記特定の値に設定させる機能と、システムシャトダウン時には、上記アドレス変換マップを不揮発性記憶に保存することにより、次回ブート時には上記アドレス変換マップをシャトダウン完了時点の状態に復元させる機能とがプログラムされ記録される請求項１５記載のコンピュータ読み取り可能な記憶媒体。

【請求項１９】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、指定された論理ブロックに関して、現在の論理ブロックと物理ブロックの組を差分登録機構に登録させる機能と、無効ブロック管理モジュールから新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとして上記アドレス変換マップに登録させる機能と、登録を終えた後この物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項２０】　チェックポイント処理において、上記差分記録機構に登録されている全ての物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記差分登録機構に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消させる機能と、リカバリ処理において、上記差分記録機構に登録されてある全ての論理ブロックに関して、上記アドレス変換マップ上でその論理ブロックにマッピングされている物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記アドレス変換マップに対して、上記差分記録機構に登録されている論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点のアドレス変換マップの内容を復元させる機能とがプログラムされ記憶される請求項１９記載のコンピュータ読み取り可能な記憶媒体。

【請求項２１】　定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、指定された論理ブロックが差分記録機構に登録されていなければ、指定された論理ブロックに関して現在の論理ブロックと物理ブロックの組を上記差分記録機構に登録させる機能と、無効ブロック管理モジュールから新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとして上記アドレス変換マップに登録させる機能と、登録を終えた後、この論理ブロックにマッピングされた物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項２２】　チェックポイント処理において、上記差分記録機構に登録されている全ての物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記差分登録機構に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消させる機能と、リカバリ処理において、上記差分記録機構に登録されてある全ての論理ブロックに関して、上記アドレス変換マップ上でその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記アドレス変換マップに対して、上記差分記録機機構に登録している論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点の上記アドレス変換マップの内容を復元する機能とがプログラムされ記憶される請求項２１記載のコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【０００１】
【発明の属する技術分野】本発明は計算機システムにおける入出力制御方式、特に高信頼計算機システムに用いて好適な入出力制御装置、及び同システムにおける入出力制御方法並びに同方法がプログラムされ記憶された計算機読取り可能な記憶媒体に関する。
【０００２】
【従来の技術】高信頼性計算機システムでは、定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には、直前に採取したチェックポイント時点にシステムの状態を戻し（ロールバック）、処理を再実行することによって、システムの一時故障を回避するようなチェックポイント・ロールバック機構を持つものが考えられていた。このチェックポイント・ロールバック機構を内蔵し、かつ、二次記憶装置との間でデータの格納、読み出しを行うような計算機システムにおいては、障害発生によるロールバックの後でも、主記憶やプロセッサ等の状態とディスク装置の内容等入出力装置との整合性を保証する必要がある。
【０００３】このため、通常チェックポイント・ロールバック方式においては、オペレーティングシステム（ＯＳ）等によって要求された入出力リクエストは、次回のチェックポイントによって発行が確定した後にディスク装置に対して発行するものであった。
【０００４】
【発明が解決しようとする課題】しかしながら上述した入出力制御方法においては、少なくとも入出力装置の状態を変更する（書き込む）ような要求に関しては、次回のチェックポイント採取処理以降に発行を遅延することになっていた。更に、次回のチェックポイント処理時以降にチェックポイント時点まで遅延した要求をまとめて発行するため、チェックポイント・ロールバック方式によって高信頼化をはかる計算機システムの性能を劣化する一つの原因となっていた。
【０００５】これを解消するために、二次記憶装置上で一つの論理ブロックに対して必ず二つの物理ブロックを割り当てることにより、最新チェックポイント時点の論理ブロックのイメージを残したまま、もう一つの物理ブロックに新たに発生した書き込み処理を行なう方式が考えられていた。この方式によれば、チェックポイントを待たずに二次記憶装置に対するライト処理を行なうことができる。しかしながら、この場合、実際の物理ディスク容量の５０％のデータ領域が有効なディスク容量として使用できないため、物理ディスクの容量効率が悪いといった問題があった。
【０００６】一方、二次記憶装置に対して書き込み要求が発生した際には、二次記憶装置に書き込むデータを一旦不揮発性メモリ上に置き、次回のチェックポイントを経た後にこの不揮発性メモリ上のデータを二次記憶装置に書き込み、同じくチェックポイントを待たずに二次記憶装置に対するライト処理を行なう方法が特願平７－１５１７３７号で提案されている。しかしながら、このことを実現するためには、不揮発性メモリを含めその周辺部分に特別なハードウェアが必要であるため、実現が容易ではないといった問題があった。
【０００７】そこで本発明は上記事情を考慮してなされたものであり、実現が容易で性能劣化が小さく、かつ二次記憶装置の容量効率を改善することのできる上述不具合を解消した計算機システムにおける入出力制御装置、及び同システムにおける入出力制御方法並びに同方法がプログラムされ記録されたコンピュータ読取り可能な記憶媒体を提供することを目的としている。
【０００８】
【課題を解決するための手段】本発明の計算機システムにおける入出力制御方式は、定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、上記二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも一つの物理ブロックを割り当て、各論理ブロックに対して、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当てるアドレス変換マップと、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールを有し、入出力モジュールが二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、入出力モジュールが二次記憶装置に対する読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージを持つ物理ブロックとし、リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、チェックポイント時点の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージをもつ物理ブロックとすることを特徴とする。
【０００９】また、アドレス変換マップには、更にそれぞれの論理ブロックに最後に書き込みを行った時刻を記録し、上記入出力モジュールが二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対し

て最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして登録すると共にこの物理ブロックに書き込み処理を行なった時刻をアドレス変換マップに登録し、入出力モジュールが二次記憶装置からの読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、リカバリ時には、アドレス変換マップ上の各論理ブロックに対して、チェックポイントタイムスタンプよりも新しいタイムスタンプを持つ最新の物理ブロックへのポインタが登録されている場合は最新の物理ブロックを無効ブロック管理モジュールに渡して解放し、チェックポイント時点の物理ブロックへのポインタを最新の物理ブロックへのポインタとしてアドレス変換マップに登録することも特徴とする。

【００１０】更に、最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関して、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構を有し、入出力モジュールが二次記憶装置への書き込み要求を受け付けたときには、この時点のこの論理ブロックと物理ブロックのアドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行い、チェックポイント処理では、上記差分記録機構に記録した全ての物理ブロックを無効ブロック管理モジュールに解放し、上記差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、アドレス変換マップでその論理ブロックにマップされている物理ブロックを無効ブロック管理モジュールに渡して解放し、アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点のアドレス変換マップを復元することも特徴とする。

【００１１】また、入出力モジュールが二次記憶装置への書き込み要求を受け付けたときに、もし指定された論理ブロックが差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックのアドレス変換マップ上のマッピングを差分記録機構に記録し、無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行い、もし指定きれた論理ブロックが差分記録機構に登録されていればアドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、チェックポイント処理では、上記差分記録機構に記録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、アドレス変換マップでその論理ブロックにマップされている物理ブロックを無効ブロック管理モジュールに渡して解放し、アドレス変換マップに対して、差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点のアドレス変換マップの内容を復元することも特徴とする。

【００１２】上記構成によれば、チェックポイント・ロールバック方式の計算機システムで、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上する。更に、二次記憶装置上で冗長な物理ブロックの数も減少できるため、二次記憶装置の容量効率も改善できる。

【００１３】

【発明の実施の形態】以下に本発明の実施形態を図面を参照して説明する。

（第１実施形態）図１は本発明の第１実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、１１は計算機本体であり、本発明と関係するところでは、アドレス変換マップ１１１と、ディスク入出力モジュール（ディスクＩ／Ｏモジュール）１１２と、無効ブロック管理モジュール１１３がメモリに割り付けられ格納される。１２は二次記憶装置となる大容量のディスク装置、１３は不揮発性メモリ（ＮＶＲＡＭ）である。

【００１４】通常、ディスク入出カモジュール１１２は、ファイルシステムからディスク入出力要求を受け取り、ディスク装置１２へ入出力要求を発行する。本実施形態によれば、ディスク入出力モジュール１１２でディスク装置１２のデータブロックの論理アドレスと物理アドレスを管理する。

【００１５】アドレス変換マップ１１１の構造を図２に示す。アドレス変換マップ１１１は、ディスク装置１２の全ての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持つ。

【００１６】ここで、有効物理ブロックとは、この論理ブロックの最新のディスクイメージを持つ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージを持ち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。

【００１７】全ての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ１１１の有効物理ブロックアドレスエントリには、この物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは旧有効物理ブロックエントリにはＮＵＬＬを記録する。

【００１８】図３は、無効ブロック管理モジュール１１３の内部構造を示す。無効ブロック管理モジュール１１３は、物理ブロックのフリーリストを管理する。即ち、ディスク装置１２上の全ての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでもない物理ブロック（無効ブロックアドレスリストＢ、Ｃ、…他）を管理するモジュールである。

【００１９】以下に、図４に示すディスク入出力モジュール１１２が書き込み要求を受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【００２０】ディスク入出力モジュール１１２が書き込み要求を受け取ったとき、ディスク入出力モジュール１１２は、ステップＳ４１にてアドレス変換マップ１１１を用いて、論理アドレスに対して旧有効アドレスが存在するか否かチェックする。アドレス変換マップ１１１で、指定された論理ブロックに対して旧有効な物理ブロックアドレスが登録されている場合には（ステップＳ４１のＹ）、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう（ステップＳ４２）。一方、アドレス変換マップ１１１で、指定された論理ブロックに対して旧有効な論理ブロックがＮＵＬＬであった場合には（ステップＳ４１のＮ）、指定された論理ブロックに対して、現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録（ステップＳ４３）し、無効ブロック管理モジュール１１３から新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ１１１に登録する（ステップＳ４４）。そして、この物理ブロックに対して書き込み処理（ステップＳ４２）を行なう。

【００２１】ところで、チェックポイント処理では、アドレス変換マップ１１１に登録されている全ての旧有効物理ブロックを無効ブロック管理モジュール１１３に渡すことにより解放し、全ての旧有効物理ブロックエントリの値をＮＵＬＬにする。また、リカバリ処理では、アドレス変換マップ１１１において、旧有効物理アドレスエントリの値がＮＵＬＬでない全ての論理ブロックに対して、有効な物理ブロックを無効ブロック管理モジュール１１３に渡すことで解放し、旧有効物理ブロックアドレスエントリの値を有効物理ブロックアドレスエントリに移動し、旧有効物理ブロックアドレスエントリの値はＮＵＬＬとする。更に、システムシャトダウン時には、アドレス変換マップ１１１を不揮発性メモリ（ＮＶＲＡＭ）１３に保存することにより、次回のブート時にはアドレス変換マップ１１１をシャトダウン完了時点の状態に復元するものである。

（第２実施形態）図５は本発明の第２実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、２１は計算機本体であり、本実施形態と関係するところでは、アドレス変換マップ２１１と、ディスク入出力モジュール２１２（ディスクＩ／Ｏモジュール）と、無効ブロック管理モジュール２１３、タイムスタンプ機能を有すタイムスタンプ２１４がメモリに割り付けられ格納される。２２は二次記憶装置となる大容量のディスク装置、２３は不揮発性メモリ（ＮＶＲＡＭ）である。

【００２２】ここで示す実施形態では、ディスク入出力モジュール２１２にて、ディスク装置２２のデータブロックの論理アドレスと物理アドレス、および物理ブロックに書き込み処理を行なった時刻を管理する。

【００２３】図６は、本実施形態のディスク入出力モジュール２１２で用いるアドレス変換マップ２１１の例を示す図である。アドレス変換マップ２１１では、ディスク装置２２の全ての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持ち、更に、ＮＵＬＬでない物理ブロックアドレスエントリにはその物理ブロックに最後に書き込み処理を行なった時刻を示すタイムスタンプ２１５を持つ。

【００２４】ここで、有効物理ブロックとは、この論理ブロックの最新のディスクイメージをもつ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージをもち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。

【００２５】全ての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ２１１の有効物理ブロックアドレスエントリにはこの物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは旧有効物理ブロックエントリにはＮＵＬＬを記録する。

【００２６】図７に無効ブロック管理モジュール２１３の構造を示す。無効ブロック管理モジュール２１３は、図１に示す無効ブロック管理モジュール１１３同様、物理ブロックのフリーリストを管理する。つまり、ディスク装置２２上の全ての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでも

ない物理ブロックを管理するモジュールである。ここでは更に、無効ブロック管理モジュール２１３は定期的にアドレス変換マップ２１１を走査し、有効物理ブロックアドレスエントリのタイムスタンプ２１５が最新チェックポイント時点のタイムスタンプ（チェックポイントタイムスタンプ）２１４よりも古い場合には、この有効物理ブロックと同じ論理ブロックに割り当てられている旧有効物理ブロックを解放し、この旧有効物理ブロックエントリの値をＮＵＬＬにする。

【００２７】以下、図８に示すディスク入出力モジュール２１２がライトリクエストを受け取ったときの動作のフローチャートを参照して、本実施形態の動作について説明する。

【００２８】ディスク入出力モジュール２１２が書き込み要求を受け取ったときに、ディスク入出力モジュール２１２は、まず、有効ブロックのタイムスタンプが最新チェックポイントのタイムスタンプ２１４より古いか否かをチェックする（ステップＳ８１）。アドレス変換マップ２１１で、指定された論理ブロックに対して有効な物理ブロックのタイムスタンプ２１５がチェックポイントタイムスタンプ２１４よりも古くない場合には（ステップＳ８１のＮ）、ステップＳ８２の処理に移り、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう。アドレス変換マップ２１１で、指定された論理ブロックに対して有効な物理ブロックのタイムスタンプ２１５がチェックポイントタイムスタンプ２１４よりも古い場合には（Ｓ８１のＹ）、更にステップＳ８３において旧有効アドレスが存在するか否かをチェックする。指定された論理ブロックの旧有効物理ブロックアドレスがＮＵＬＬでなければ（ステップＳ８３のＹ）、この旧有効物理ブロックを無効ブロック管理モジュール２１３に渡すことで解放し（ステップＳ８５）、指定された論理ブロックに対して、現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録し（ステップＳ８４）、無効ブロック管理モジュール２１３から新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ２１１に登録する（ステップＳ８６）。そして、この物理ブロックに対して書き込み処理を行なう（ステップＳ８２）。

【００２９】ところで、チェックポイント処理では、最新チェックポイントの時刻をチェックポイントタイムスタンプ２１４として登録する。リカバリ処理では、アドレス変換マップ２１１上の全ての論理ブロックに対して、もし、その論理ブロックに対応する有効物理ブロックのタイムスタンプ２１５がチェックポイントタイムスタンプ２１４よりも新しい場合は、有効物理ブロックを無効ブロック管理モジュール２１３に渡すことで解放し、旧有効物理ブロックアドレスエントリの値を有効物理ブロックアドレスに移動し、旧有効物理ブロックアドレスエントリの値をＮＵＬＬにする。システムシャトダウン時には、アドレス変換マップ２１１を不揮発性メモリ（ＮＶＲＡＭ）２３に保存することにより、次回のブート時にはアドレス変換マップ２１１をシャトダウン完了時点の状態に復元するものである。

（第３実施形態）図９は本発明の第３実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、３１は計算機本体であり、本実施形態と関係するところでは、アドレス変換マップ３１１と、ディスク入出力モジュール（ディスクＩ／Ｏモジュール）３１２と無効ブロック管理モジュール３１３、差分記録機構３１５がメモリに割り付けられ格納される。３２は二次記憶装置となる大容量のディスク装置、３３は不揮発性メモリ（ＮＶＲＡＭ）である。

【００３０】本実施形態では、最新チェックポイント以降に新たに書き込み処理を行なった論理ブロックについて、チェックポイント時点の論理ブロックイメージをもつ物理ブロックを、差分記録機構３１５によって管理するものである。

【００３１】図１０に、ディスク入出力モジュール３１２で用いる差分記録機構３１５を示す。差分記録機構３１５は、最新チェックポイント以降に新たに書き込み処理を行なったディスク装置３２の論理ブロックに関して、論理ブロックアドレスとチェックポイント時点で割り当てれられていた物理ブロックのアドレスの組を記録する。

【００３２】図１１は、無効ブロック管理モジュール３１３の構造を示す図である。無効ブロック管理モジュール３１３は、物理ブロックのフリーリストを管理する。つまり、ディスク記憶装置３２上の全ての物理ブロックのうち、どの論理ブロックにも割り当てられていない物理ブロックを管理するモジュールである。

【００３３】以下、図１２のディスク入出力モジュール３１２が書き込み要求を受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【００３４】まず、ディスク入出力モジュール３１２が書き込み要求を受け取ったときに、ディスク入出力モジュール３１２は、指定された論理ブロックに関して、現在の論理ブロックと物理ブロックの組を差分記録機構３１５に登録し（ステップＳ１２１）、無効ブロック管理モジュール３１３から新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとしてアドレス変換マップ３１１に登録する（ステップＳ１２２）。この後に、この物理ブロックに対して書き込み処理（ステップＳ１２３：アドレス変換マップ３１１で指定した物理ブロックにデータライト）を行なう。

【００３５】ところで、チェックポイント処理では、差分記録機構３１５に登録されている全ての物理ブロック

を無効ブロック管理モジュール３１３に渡すことにより解放し、差分記録機構３１５に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消する。リカバリ処理では、差分記録機構３１５に登録してある全ての論理ブロックに関して、アドレス変換マップ３１１上でその論理ブロックにマッピングされている物理ブロックを無効ブロック管理モジュール３１３に渡すことにより解放し、アドレス変換マップ３１１に対して、差分記録機構３１５に登録している論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点のアドレス変換マップ３１１を復元する。尚、アドレス変換マップ３１１、無効ブロック管理モジュール３１３、差分記録機構３１５を、ロールバックするメモリ上に構成した場合には、各構造の状態はロールバックによって自動的にチェックポイント時点の状態を復元できるため、ここでのリカバリ処理は必要がなくなる。

【００３６】また、システムシャットダウン時には、一対一のアドレス変換マップをもつ通常のシステムと同様に、アドレス変換マップ３１１を不揮発性メモリ３３に保存することにより、次回のブート時にはアドレス変換マップをシャトダウン完了時点の状態に復元する。

（第４実施形態）図１３は本発明の第４実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、４１は計算機本体であり、本発明実施形態と関係するところでは、アドレス変換マップ４１１と、ディスク入出力モジュール（ディスクＩ／Ｏモジュール）４１２と、無効ブロック管理モジュール４１３と、差分記録機構４１５がメモリに割り付けられ格納される。４２は二次記憶装置となる大容量のディスク装置、４３は不揮発性メモリ（ＮＶＲＡＭ）である。

【００３７】ここで示す実施形態では、最新チェックポイント以降に新たに書き込み処理を行なった論理ブロックについて、チェックポイント時点の論理ブロックイメージをもつ物理ブロックを、差分記録機構４１５によって管理する。差分記録機構４１５は、最新チェックポイント以降に新たに書き込み処理を行なったディスク装置４２の論理ブロックに関して、論理ブロックアドレスとチェックポイント時点で割り当てれられていた物理ブロックのアドレスの組を記録する（図１４参照）。

【００３８】図１５に無効ブロック管理モジュール４１３の構造を示す。無効ブロック管理モジュール４１３は、物理ブロックのフリーリストを管理する。つまり、ディスク装置４２上の全ての物理ブロックのうち、どの論理ブロックにも割り当てられていない物理ブロックを管理するモジュールである。

【００３９】以下、図１６のディスク入出力モジュール４１２が書き込み要求を受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【００４０】ディスク入出力モジュール４１２が書き込み要求を受け取ったときに、ディスク入出力モジュール４１２は、まず、その論理アドレスが差分記録機構４１５に存在するか否かをチェックする（ステップＳ１６１）。もし、指定された論理ブロックが差分記録機構４１５に登録されていなければ（Ｓ１６１のＮ）、指定された論理ブロックに関して、現在の論理ブロックと物理ブロックの組を差分記録機構４１５に登録し（ステップＳ１６３）、無効ブロック管理モジュール４１３から新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとしてアドレス変換マップ４１１に登録する（ステップＳ１６４）。この後に、この論理ブロックにマッピングされた物理ブロックに対して書き込み処理を行なう（ステップＳ１６２）。一方、差分記録機構４１５に存在するときは（ステップＳ１６１のＹ）、ステップＳ１６２に移行してアドレス変換マップ４１１により指定された物理ブロックに対して書き込み処理を行う。

【００４１】ところで、チェックポイント処理では、差分記録機構４１５に登録されている全ての物理ブロックを無効ブロック管理モジュール４１３に渡すことにより解放し、差分記録機構４１５に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消する。リカバリ処理では、差分記録機構４１５に登録してある全ての論理ブロックに関して、アドレス変換マップ４１１上でその論理ブロックにマップされている物理ブロックを無効ブロック管理モジュール４１３に渡すことにより解放し、アドレス変換マップ４１１に対して、差分記録機構４１５に登録している論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点のアドレス変換マップ４１１の内容を復元する。尚、図９に示す実施形態同様、アドレス変換マップ４１１、無効ブロック管理モジュール４１３、差分記録機構４１５をロールバックするメモリ上に構成した場合には、各構造の状態はロールバックによって自動的にチェックポイント時点の状態を復元できるため、リカバリ処理は必要なくなる。システムシャトダウン時には、一対一のアドレス変換マップをもつ通常のシステムと同様に、アドレス変換マップ４１１を不揮発性メモリ４３に保存することにより、次回のブート時にはアドレス変換マップ４１１をシャットダウン完了時点の状態に復元する。このことにより、チェックポイント・ロールバック方式の計算機システムで、二次記憶装置に対する入出力を遅延なしで発行することにより、システムのパフォーマンスが向上し、かつ、二次記憶装置の容量効率も改善する。

（第５実施形態）図１７は、本発明の第５実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、５１は計算機本体であり、本発明実施形態と関係するところでは、アドレス変換マップ５１１

と、ディスク入出力モジュール（ディスクＩ／Ｏモジュール）５１２、無効ブロック管理モジュール５１３、アドレス変換マップページング機構５１６がメモリに割り付けられ格納される。５２は二次記憶装置となる大容量のディスク装置であり、アドレス変換マップ５２１を持つ。５３は不揮発性メモリ（ＮＶＲＡＭ）である。

【００４２】ここで示す実施形態では、通常、ディスク入出力モジュール５１２はファイルシステムからディスク入出力リクエストを受け取り、ディスク装置５２へ入出力リクエストを発行する。また、ディスク入出力モジュール５１２にてディスク装置５２のデータブロックの論理アドレスと物理アドレスを管理する。

【００４３】図１８にディスク入出力モジュール５１２で用いるアドレス変換マップの例を示す。アドレス変換マップ５１１ではディスク装置５２の全ての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持つ。つまり、アドレス変換マップ５１１の各エントリのうち、メインメモリ上には有限個のエントリだけが存在し、その他のエントリはディスク装置５２上に存在する。

【００４４】ここで、アドレス変換マップ５１１において、有効物理ブロックとは、この論理ブロックの最新のディスクイメージをもつ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージをもち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。全ての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ５１１の有効物理ブロックアドレスエントリには、この物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは、旧有効物理ブロックエントリにはＮＵＬＬを記録する。

【００４５】図１９に無効ブロック管理モジュール５１３の構造を示す。無効ブロック管理モジュール５１３は、物理ブロックのフリーリストを管理する。つまり、ディスク装置５２上の全ての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでもない物理ブロックを管理するモジュールである。

【００４６】以下、図２０のディスク入出力モジュール５１２がライトリクエストを受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【００４７】ディスク入出力モジュール５１２が書き込みリクエストを受け取ったときには、ディスク入出力モジュール５１２は以下の手順でライトリクエストを発行する。まず、指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ５１１に存在しない場合は（ステップＳ５０２のＮ）、アドレス変換マップページング機構５１６がアドレス変換マップ５１１のページングを行い、指定の論理ブロックに対応するアドレス変換マップ５１１のエントリをメモリ上に置くものである（ステップＳ５０８、Ｓ５１０）。指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ５１１に存在し（ステップＳ５０２のＹ）、アドレス変換マップ５１１にて指定された論理ブロックに対して旧有効な物理ブロックアドレスが登録されている場合には（Ｓ５０４のＹ）、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう（ステップＳ５０６）。アドレス変換マップ５１１で、指定された論理ブロックに対して旧有効な物理ブロックがＮＵＬＬであった場合には（ステップＳ５０４のＮ）、指定された論理ブロックに対して、現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録し（ステップＳ５１２）、無効ブロック管理モジュール５１３から新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ５１１に登録する（ステップＳ５１４）。そして、この物理ブロックに対して書き込み処理を行なう（ステップＳ５０６）。

【００４８】システムシャトダウン時には、アドレス変換マップ５１１を不揮発性メモリ５３に保存することにより、次回のブート時にはアドレス変換マップをシャットダウン完了時点の状態に復元するものである。

【００４９】ところで、アドレス変換マップページング機構５１６の機能を述べるに、メモリ上のアドレス変換マップ５１１はロールバックで最新チェックポイント時点の状態に戻るものである。

【００５０】ディスク装置５２上のある論理ブロックに対してアドレス変換マップ５１１を書き換える必要がある場合、もし、その論理ブロックに対するアドレス変換マップ５１１のエントリがメモリ上に存在するときは、そのエントリの各有効／旧有効の値を適切に書きかえるだけでよいものである。一方、エントリがメモリ上に存在しない、即ちディスク上にページアウトされているときは、適当なアルゴリズム（ＬＲＵ等）でメモリ上のアドレス変換マップ５１１の一つ以上のエントリを選択し、この選択したエントリをディスクへページアウトする。このとき、ページアウト先の論理ブロックに対するアドレス変換マップ５１１のエントリも正しく設定する。斯様な手順にてメモリ上のアドレス変換マップ５１１の一部を置き換えるものである。

（第６実施形態）図２１は、本発明の第６実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、６１は計算機本体であり、本発明実施形態と関係するところでは、アドレス変換マップ６１１

と、ディスク入出力モジュール（ディスクＩ／Ｏモジュール）６１２、無効ブロック管理モジュール６１３、アドレス変換マップページング機構６１６、アドレス差分テーブル６１７がメモリに割り付けられ格納される。６２は二次記憶装置となる大容量のディスク装置であり、アドレス変換マップ６２１を持つ。６３は不揮発性メモリ（ＮＶＲＡＭ）である。

【００５１】ここで示す実施形態では、通常、ディスク入出力モジュール６１２はファイルシステムからディスク入出力リクエストを受け取り、ディスク装置６２へ入出力リクエストを発行する。本実施形態では、ディスク入出力モジュール６１２でディスク装置６２のデータブロックの論理アドレスと物理アドレスを管理する。

【００５２】図２２にディスク入出力モジュール６１２で用いるアドレス変換マップ６１１の例を示す。アドレス変換マップ６１１では、ディスク装置６２のすべての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持つ。アドレス変換マップ６１１の各エントリのうち、ロールバックによって最新のチェックポイントの状態に戻らない記憶領域（例えばＰＣＩデバイス上のメモリ）上に有限個のエントリだけが存在し、その他のエントリはディスク装置６２上に存在する。

【００５３】アドレス変換マップ６１１において、有効物理ブロックとは、この論理ブロックの最新のディスクイメージをもつ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージをもち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。すべての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ６１１の有効物理ブロックアドレスエントリにはこの物理ブロックのアドレスを格納する。

【００５４】旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは旧有効物理ブロックエントリにはＮＵＬＬを記録するものである。

【００５５】図２３に無効ブロック管理モジュール６１３の構成を示す。無効ブロック管理モジュール６１３は、物理ブロックのフリーリストを管理する。つまり、ディスク装置６２上のすべての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでもない物理ブロックを管理するモジュールである。

【００５６】図２４にアドレスマップ差分テーブル６１７を示す。アドレスマップ差分テーブル６１７は、最新チェックポイント時点以降、メモリ上のアドレス変換マップ６１１の変更履歴を保存している。メモリ上のアドレス変換マップ６１１に割り当てられた領域の内容が、アドレス変換マップ６１１のページングによってディスク装置６２に書き出す際、ページアウトされるエントリの内容をアドレスマップ差分テーブル６１７に保存する。メモリ上のアドレス変換マップ６１１のエントリの内容をアドレスマップ差分テーブル６１７に保存し、ディスク装置６２に書き出した後に、このメモリ上のアドレス変換マップ６１１のエントリは無効になる。新たに別のエントリをディスク装置６２上のアドレス変換マップ６１２からメモリ上のアドレス変換マップ６１１に読み込む際には、メモリ上のアドレス変換マップ６１１の無効な領域にのみエントリを読み込むことができる。

【００５７】以下、図２５に示すディスク入出力モジュール６１２がライトリクエストを受け取ったときの動作を示すフローチャートを参照して、本実施形態の動作を説明する。

【００５８】ディスク入出力モジュール６１２が書き込みリクエストを受け取ったときには、ディスク入出力モジュール６１２は以下の手順でライトリクエストを発行する。まず、指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ６１１に存在しない場合は（ステップＳ６０２のＮ）、アドレス変換マップページング機構６１６がアドレス変換マップ６１１のページングを行ない、指定の論理ブロックに対応するアドレス変換マップ６１１のエントリをメモリ上に置き（ステップＳ６１０、Ｓ６１２）、ステップＳ６０４に移行する。指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ６１１に存在し（ステップＳ６０２のＹ）、アドレス変換マップ６１１にて指定された論理ブロックに対して旧有効な物理ブロックアドレスが登録されている場合には（ステップＳ６０４のＹ）、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう（ステップＳ６０８）。一方、アドレス変換マップ６１１で、指定された論理ブロックに対して旧有効な物理ブロックがＮＵＬＬであった場合には（ステップＳ６０４のＮ）、指定された論理ブロックに対して現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録し（ステップＳ６１４）、無効ブロック管理モジュール６１３から新たに物理ブロックを確保すると共に、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ６１１に登録し（ステップＳ６１４）、この物理ブロックに対して書き込み処理を行なう（ステップＳ６１６）。

【００５９】チェックポイント処理では、アドレス変換マップ６１１に登録しているすべての旧有効物理ブロックを無効ブロック管理モジュールに渡すことにより解放し、すべての旧有効物理ブロックエントリの値をＮＵＬＬにする。また、リカバリ処理では、アドレス変換マッ

プ６１１において、旧有効物理アドレスエントリの値がＮＵＬＬでないすべての論理ブロックに対して、有効な物理ブロックを無効ブロック管理モジュールに渡すことで解放して、旧有効物理ブロックアドレスエントリの値を有効物理ブロックアドレスエントリに移動し、旧有効物理ブロックアドレスエントリの値はＮＵＬＬとする。システムシャトダウン時には、アドレス変換マップ６１１を不揮発性メモリ６３に保存することにより、次回のブート時にはアドレス変換マップ６１１をシャトダウン完了時点の状態に復元する。

【００６０】ところで、アドレス変換マップページング機構６１６について述べるに、メモリ上のアドレス変換マップ６１１はロールバックで最新チェックポイント時点の状態に戻らないものである。

【００６１】ディスク装置６２上のある論理ブロックに対してアドレス変換マップ６１１を書き換える必要がある場合、もし、その論理ブロックに対するアドレス変換マップのエントリがメモリ上に存在するときは、そのエントリの各有効／旧有効の値を適切に書きかえるだけでよい。

【００６２】もし、エントリがメモリ上に存在しない（ディスク上にページアウトされている）ときは、適当なアルゴリズム（ＬＲＵ等）でメモリ上のアドレス変換マップ６１１の一つ以上のエントリを選択し、この選択したエントリをメモリ上のアドレスマップ差分テーブル６１７にコピーする。また、選択したエントリをディスク装置６２へページアウトする。このとき、ページアウト先の論理ブロックに対するアドレス変換マップのエントリも正しく設定する。斯様な手順にてメモリ上のアドレス変換マップ６１１の一部を置きかえるものである。

【００６３】チェックポイント処理の一環で、アドレスマップ差分テーブル６１７をクリアする。リカバリ処理の一環では、アドレスマップ差分テーブル６１７に記録した各エントリを、アドレス変換マップ６１１に書き戻すものである。

【００６４】尚、上記各実施形態にて述べた手法は、当該手法が計算機にて実行可能となるようプログラムされたコンピュータ読取り可能な記憶媒体として提供することは勿論可能である。

【００６５】

【発明の効果】以上詳記したように本発明によれば、チェックポイント・ロールバック方式の計算機システムで、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上する。更に、二次記憶装置上で冗長な物理ブロックの数も減少できるため、二次記憶装置の容量効率も改善できる。更にタイムスタンプを用いることにより、チェックポイント処理が軽くなり一層システムのパフォーマンスが向上する。

【００６６】また、図１と図５に示す実施形態が論理ブロックと物理ブロックが１対多数の特殊なマッピングに対して適用できるのに対して、図９と図１３に示す実施形態は１対１の一般的なマッピングに対して適用でき、チェックポイント・ロールバック方式の計算機システムにて、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上する。更に、二次記憶装置上で冗長な物理ブロックの数も減少できるため、二次記憶装置の容量効率も改善できるものである。

【００６７】また、図１３に示す実施形態では図９に示す実施形態に比較してチェックポイント時の処理が軽くなるため、システムのパフォーマンスが更に向上する。

【００６８】更に、本発明によれば、チェックポイント/ロールバック方式の計算機システムで、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上できる。加えて、二次記憶装置の容量にかかわらずメインメモリ上やのアドレス変換マップで必要な記憶領域が一定であり、メインメモリを圧迫することはない。チェックポイント処理の負荷も軽減される。

【図面の簡単な説明】

【図１】本発明の第１実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図２】同実施形態に係わるアドレス変換マップの構造を示す図。

【図３】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

【図４】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図５】本発明の第２実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図６】同実施形態に係わり、アドレス変換マップの構造を示す図。

【図７】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

【図８】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図９】本発明の第３実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図１０】同実施形態に係わる差分記録機構の構造を示す図。

【図１１】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

【図１２】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図１３】本発明の第４実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

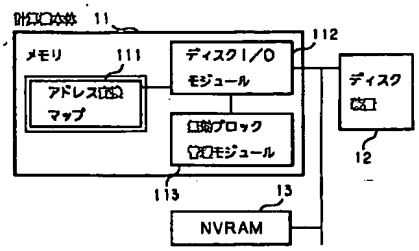
【図１４】同実施形態に係わる差分記録機構の構造を示す図。

【図１５】同実施形態に係わる無効ブロック管理モジュ

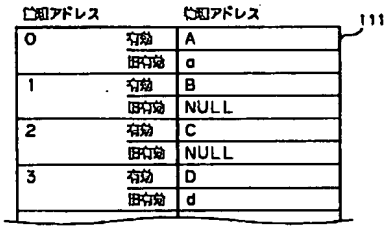
ールの構造を示す図。
【図１６】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。
【図１７】本発明の第５実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。
【図１８】同実施形態に係わるドレス変換マップの構造を示す図。
【図１９】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。
【図２０】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。
【図２１】本発明の第６実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。
【図２２】同実施形態に係わるアドレス変換マップの構造を示す図。
【図２３】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。
【図２４】同実施形態に係わるアドレス差分テーブルの構造を示す図。
【図２５】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【符号の説明】
１１（２２，３２，４２，５１，６１）…計算機本体、１２（２２，３２，４２，５２，６２）…ディスク装置、１３（２３，３３，４３，５３，６３）…不揮発性メモリ、１１１（２１１，３１１，４１１，５１１，６１１）…アドレス変換マップ、１１２（２１２，３１２，４１２，５１２，６１２）…ディスク入出力モジュール、１１３（２１３，３１３，４１３，５１３，６１３）…無効ブロック管理モジュール、２１４…タイムスタンプ、３１５（４１５）…差分記録機構、５１６（６１６）…アドレス変換マップページング機構、６１７…アドレス差分テーブル。

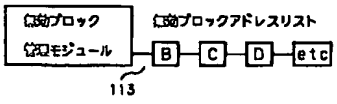
【図１】



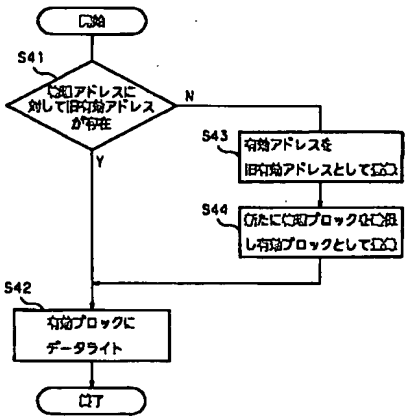
【図２】



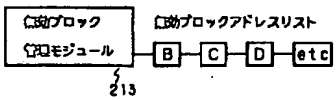
【図３】



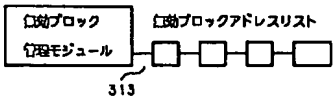
【図４】



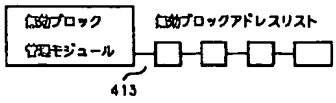
【図７】



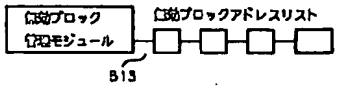
【図１１】



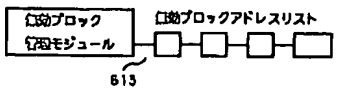
【図１５】



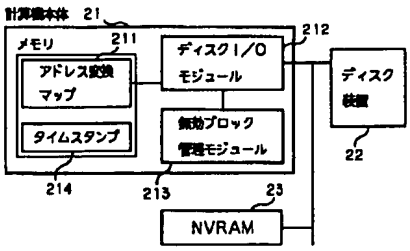
【図１９】



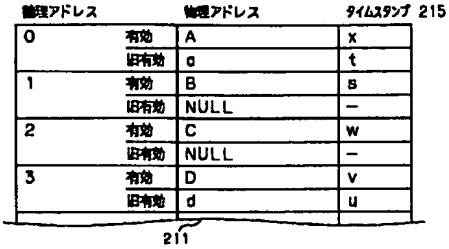
【図２３】



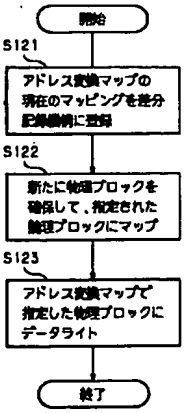
【図５】



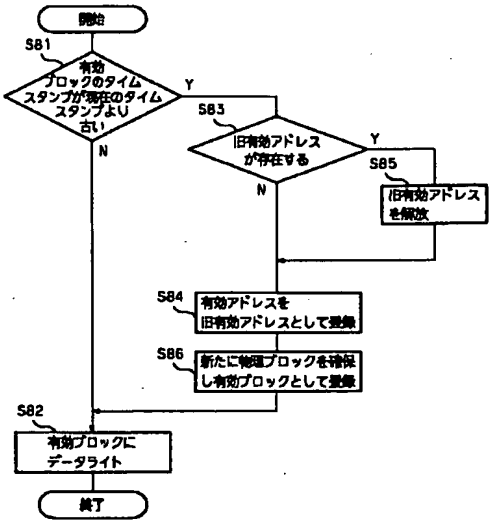
【図６】



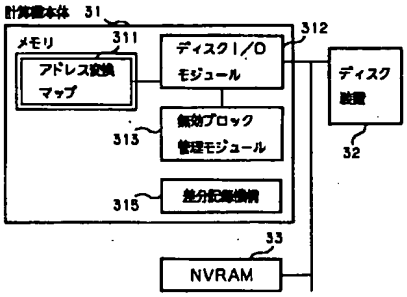
【図１２】



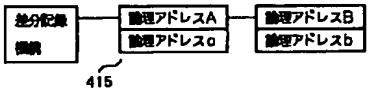
【図８】



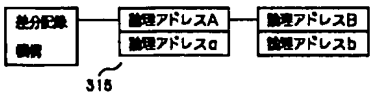
【図９】



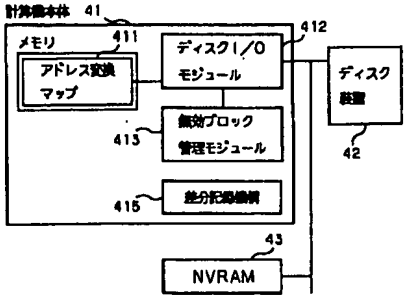
【図１４】



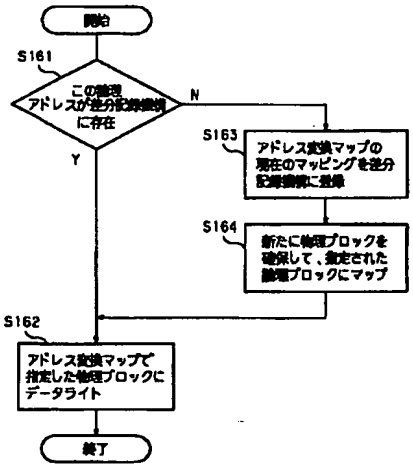
【図１０】



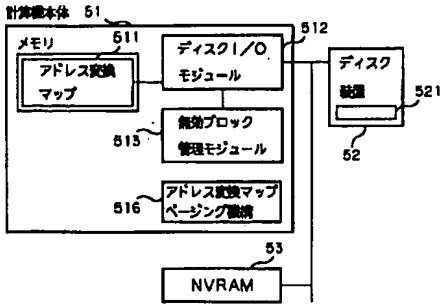
【図１３】



【図１６】



【図１７】

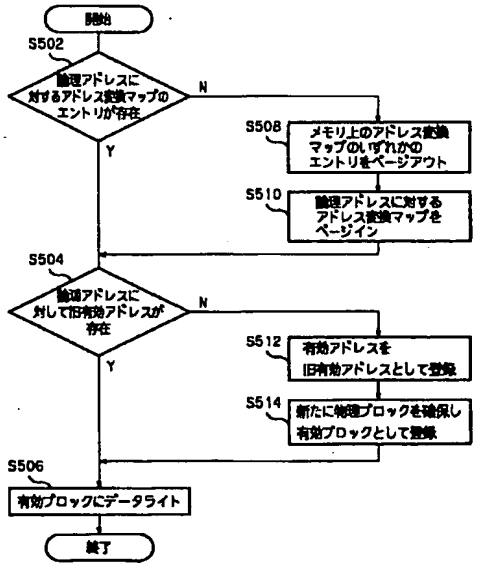


【図１８】

論理アドレス		物理アドレス
s	有効	A
	旧有効	a
t	有効	B
	旧有効	NULL
u	有効	C
	旧有効	NULL
v	有効	D
	旧有効	d

511

【図２０】

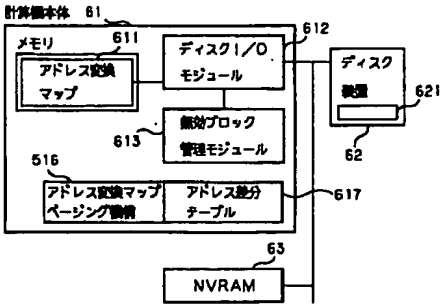


【図２２】

論理アドレス		物理アドレス
s	有効	A
	旧有効	a
t	有効	B
	旧有効	NULL
u	有効	C
	旧有効	NULL
v	有効	D
	旧有効	d

611

【図２１】



【図２４】

論理アドレス		物理アドレス
s	有効	A
	旧有効	α
t	有効	B
	旧有効	NULL
NULL	有効	NULL
	旧有効	NULL
NULL	有効	NULL
	旧有効	NULL

617

【図２５】